

AMa104

Homework #3

(Projections, Orthogonality, Least Squares, Gram-Schmidt)

Handed out: 1 November 1996

Due at noon: 8 November 1996

- **Problem 1.** (This is a warmup problem for Problem 2.) Consider the (somewhat trivial) matrix equation $Ax = b$, where A consists of a single column vector a , and where

$$a = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad b = \begin{pmatrix} 3 \\ 7 \\ 11 \end{pmatrix}.$$

We saw in class that such an overdetermined system can be solved in a certain sense by solving the normal equations:

$$A^T Ax = A^T b.$$

1. Form the normal equations and solve for x .
2. Identify and write down the least-squares orthogonal projector, $P = A(A^T A)^{-1}A^T$.
3. Verify that P is truly a projector (i.e., that it is idempotent and self-adjoint).
4. Write down the projection Ax of b , namely Pb .
5. Verify that the residual $b - Ax$ is orthogonal to $\mathcal{R}(A)$, the range space of A .
6. Consider now the following functional, and its equivalent forms, for the general case of $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, and $m > n$:

$$\begin{aligned} E(x) &= \frac{1}{2} \|b - Ax\|_{l^2}^2 = \frac{1}{2} (b - Ax, b - Ax)_{l^2} \\ &= \frac{1}{2} [(b, b)_{l^2} - (b, Ax)_{l^2} - (Ax, b)_{l^2} + (Ax, Ax)_{l^2}] \\ &= \frac{1}{2} [b^T b - b^T Ax - (Ax)^T b + (Ax)^T Ax] \\ &= \frac{1}{2} [b^T b - 2x^T A^T b + x^T (A^T A)x]. \end{aligned}$$

Show that the last equality holds; i.e., convince yourself that

$$b^T Ax = (Ax)^T b = x^T A^T b.$$

7. In the case of our simple problem, where A is a single column vector and x is a scalar, this functional reduces to a function of a single real variable, namely

$$E(x) = \frac{1}{2} [b^T b - 2(a^T b)x + (a^T a)x^2].$$

Show that for our example, the solution to the normal equations is a critical point for the above functional. Since this functional is quadratic in x , does the solution to the normal equations provide more than just a critical point?

- **Problem 2.** Consider now the general overdetermined system $Ax = b$, where $A \in \mathbb{R}^{n \times m}$, $x \in \mathbb{R}^m$, $b \in \mathbb{R}^n$, with $m > n$. Assume that the m columns a^j , $1 \leq j \leq m$, of A are linearly independent. We will think of A as being a collection of these columns:

$$A = \left(a^1 \mid a^2 \mid \cdots \mid a^m \right),$$

where $a^j \in \mathbb{R}^n$.

1. Consider again the functional:

$$E(x) = \frac{1}{2} \|b - Ax\|_2^2 = \frac{1}{2} \left[b^T b - 2x^T A^T b + x^T (A^T A)x \right].$$

Show that in this general setting, the solution to the normal equations is a critical point for the above functional, i.e., that

$$\nabla E(x) = A^T Ax - A^T b.$$

In other words, you need to show that

$$\frac{\partial E(x)}{\partial x_i} = (A^T Ax)_i - (A^T b)_i, \quad i = 1, \dots, n.$$

Some hints: It may be helpful to think about $A^T b$ as the inner-product of the columns of A with the vector b , which forms some vector that is independent of x . Similarly, $A^T A$ is just the matrix formed by the inner-product of the columns of A with themselves, again independent of x . If you write out the inner-products forming function $E(x)$ as sums, then the differentiation with respect to x_i will be straight-forward. Also, I find it useful to employ Einstein's summation convention when doing these sorts of calculations; for example, rather than writing out all of the sums in

$$\begin{aligned} E(x) &= \frac{1}{2} \left[b^T b - 2x^T A^T b + x^T (A^T A)x \right] \\ &= \frac{1}{2} \left[\sum_{i=1}^m b_i b_i - 2 \sum_{i=1}^n x_i \sum_{j=1}^m a_{ji} b_j + \sum_{i=1}^m x_i \sum_{j=1}^m \left(\sum_{k=1}^n a_{ki} a_{kj} \right) x_j \right], \end{aligned}$$

we can simply stick to the convention that all repeated indices in products are assumed to be summed, so that we can write the above simply as:

$$E(x) = \frac{1}{2} [b_i b_i - 2x_i a_{ji} b_j + x_i a_{ki} a_{kj} x_j].$$

Differentiating $E(x)$ is done term-by-term, so that we can simply differentiate the term above, and the full result is obtained by keeping the summation convention (inserting the sums back in). This simplifies the notation and clarifies things to a large degree, as you will find.

2. In the previous part you have shown that the critical points of the functional $E(x)$ satisfy the normal equations:

$$\nabla E(x) = A^T Ax - A^T b = 0.$$

Therefore, the functional $E(x)$ is (locally) maximal or minimal at the solution to the normal equations. Can you show that due to the form of $E(x)$, this is in fact a local minimum? Given that $E(x)$ is a generalized quadratic function in the components of the vector x , can you say anything about this minimum being a global minimum?

- **Problem 3.** (This is extra credit; not required! But, it is pretty useful in constructing iterative methods for linear systems...) Consider the following problem, which is very similar to Problem 2. We are given a square matrix equation $Ax = b$, where $A \in \mathbb{R}^{n \times n}$ is nonsingular, and we wish to find the “best” solution in a subspace $V \subset \mathbb{R}^n$, $V = \text{span}\{p^1, \dots, p^m\}$, where $\{p^k\}$ is a basis for V (thus, $\dim(V) = m \leq n$). We are given the fixed basis $\{p^r\}$, and are looking for a solution of the form

$$x = \sum_{r=1}^m \alpha_r p^r,$$

such that x minimizes some provided energy functional $E(x)$. Therefore, we need to derive some conditions on the free parameters $\{\alpha_r\}$ such that $E(x)$ will be minimized.

1. Consider first the least-squares energy functional:

$$\min_{x \in V} E(x), \quad \text{where } E(x) = \frac{1}{2} \|b - Ax\|_{l^2}^2.$$

Plugging the proposed x into $E(x)$ (and employing summation convention) gives

$$E(\alpha_1, \dots, \alpha_m) = \frac{1}{2} [b_i b_i - 2\alpha_r p_i^r a_{ji} b_j + \alpha_r p_i^r a_{ki} a_{kj} \alpha_s p_j^s].$$

Show that $E(\alpha_1, \dots, \alpha_m)$ is minimized at the solution to

$$Cy = d,$$

where $C_{ij} = (p^i, A^T A p^j)_{l^2}$, $d_i = (p^i, A^T b)_{l^2}$, and $y_i = \alpha_i$. In other words, show that

$$\nabla E(\alpha_1, \dots, \alpha_m) = Cy - d.$$

2. Consider now a modified energy functional:

$$E(x) = \frac{1}{2} (x, Ax)_{l^2} - (x, b)_{l^2} = \frac{1}{2} x_i a_{ij} x_j - x_i b_i.$$

We know that $E(x)$ has a critical point when x is such that $\nabla E(x) = 0$; show that this condition is:

$$\frac{1}{2} (A + A^T) x = b.$$

Now, plugging the subspace solution $x = \sum_{r=1}^m \alpha_r p^r$ into $E(x)$ gives

$$E(\alpha_1, \dots, \alpha_m) = \frac{1}{2} \alpha_r p_i^r a_{ij} \alpha_s p_j^s - \alpha_r p_i^r b_i.$$

Show that $E(\alpha_1, \dots, \alpha_m)$ has a critical point at the solution to

$$Cy = d,$$

where $C_{ij} = (p^i, \frac{1}{2}[A + A^T]p^j)_{l^2}$, $d_i = (p^i, b)_{l^2}$, and $y_i = \alpha_i$. In other words, show that

$$\nabla E(\alpha_1, \dots, \alpha_m) = Cy - d.$$

3. Assume now that the matrix A is both symmetric and positive, in which case the bilinear form (Au, v) forms an inner-product, as we proved in class. Give the simplified expressions for the results in Part 2 of this problem under these additional assumptions on A . Are the critical point conditions of Part 2 now conditions for global minima (or maxima) of $E(x)$?